Estimating changes in global temperature since the pre-industrial period

Citation for published version:
Hawkins, E, Ortega, P, Suckling, E, Schurer, A, Hegerl, G, Jones, P, Joshi, M, Osborn, TJ, Masson-delmotte, V, Mignot, J, Thorne, P & Van Oldenborgh, GJ 2017, 'Estimating changes in global temperature since the pre-industrial period', *Bulletin of the American Meteorological Society*. https://doi.org/10.1175/BAMS-D-16-0007.1

Digital Object Identifier (DOI):
10.1175/BAMS-D-16-0007.1

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Bulletin of the American Meteorological Society

The DOI for this manuscript is doi: 10.1175/BAMS-D-16-0007.1

The final published version of this manuscript will replace the preliminary version at the above DOI once it is available.

If you would like to cite this EOR in a separate work, please use the following full citation:

Hawkins, E., P. Ortega, E. Suckling, A. Schurer, G. Hegerl, P. Jones, M. Joshi, T. Osborn, V. Masson-Delmotte, J. Mignot, P. Thorne, and G. van Oldenborgh, 2017: Estimating changes in global temperature since the pre-industrial period. Bull. Amer. Meteor. Soc. doi:10.1175/BAMS-D-16-0007.1, in press.

Estimating changes in global temperature since the pre-industrial period

Ed Hawkins*, Pablo Ortega, Emma Suckling

NCAS-Climate, Department of Meteorology, University of Reading, Reading, UK

Andrew Schurer, Gabi Hegerl

School of GeoSciences, Grant Institute, University of Edinburgh, Edinburgh, UK

Phil Jones

School of Environmental Sciences, University of East Anglia, Norwich, UK and Center of Excellence for Climate Change Research, Department of Meteorology, King Abdulaziz University, Jeddah, Saudi Arabia

Manoj Joshi, Timothy J. Osborn

School of Environmental Sciences and Climatic Research Unit, University of East Anglia, Norwich, UK

Valérie Masson-Delmotte

Institut Pierre Simon Laplace, Laboratoire des Sciences du Climat et de l'Environnement (CEA-CNRS-UVSQ), Gif-sur-Yvette, France

Juliette Mignot

Climate and Environmental Physics, Physics Institute & Oeschger Center for Climate Change Research, University of Bern, Switzerland and LOCEAN/IPSL (Sorbonne Universités, UPMC-CNRS-IRD-MNHN), France

Peter Thorne

Department of Geography, Maynooth University, Maynooth, County Kildare, Ireland

Geert Jan van Oldenborgh

Koninklijk Nederlands Meteorologisch Instituut, De Bilt, Netherlands

*Corresponding author address: Ed Hawkins, Department of Meteorology, University of Reading, Reading, UK. RG6 6BB.

E-mail: e.hawkins@reading.ac.uk

ABSTRACT

The United Nations Framework Convention on Climate Change (UNFCCC) process agreed in Paris to limit global surface temperature rise to 'well below 2°C above pre-industrial levels'. But what period is 'pre-industrial'? Somewhat remarkably, this is not defined within the UNFCCC's many agreements and protocols. Nor is it defined in the IPCC's Fifth Assessment Report (AR5) in the evaluation of when particular temperature levels might be reached because no robust definition of the period exists. Here we discuss the important factors to consider when defining a pre-industrial period, based on estimates of historical radiative forcings and the availability of climate observations. There is no perfect period, but we suggest that 1720-1800 is the most suitable choice when discussing global temperature limits. We then estimate the change in global average temperature since pre-industrial using a range of approaches based on observations, radiative forcings, global climate model simulations and proxy evidence. Our assessment is that this pre-industrial period was likely $0.55-0.80$°C cooler than 1986-2005 and that 2015 was likely the first year in which global average temperature was more than 1°C above pre-industrial levels. We provide some recommendations for how this assessment might be improved in future and suggest that reframing temperature limits with a modern baseline would be inherently less uncertain and more policy-relevant.

Better defining (or altogether avoiding) the term 'pre-industrial' would aid interpretation of internationally agreed global temperature limits and estimation of the required constraints to avoid reaching those limits.

The basis for international negotiations on climate change has been to '*prevent dangerous anthropogenic interference with the climate system*', using the words of the United Nations Framework Convention on Climate Change (UNFCCC). The 2015 Paris COP21 Agreement[1], aims to maintain global average temperature *'well below* $2°C$ *above pre-industrial levels and to pursue efforts to limit the temperature increase to* $1.5°C$ *above pre-industrial levels'*. However, there is no formal definition of what is meant by 'pre-industrial' in the UNFCCC or the Paris Agreement. Neither did the Fifth Assessment Report (AR5) of the Intergovernmental Panel on Climate Change (IPCC) use the term when discussing when global average temperature might cross various levels, due to the lack of a robust definition (Kirtman et al., 2013).

Ideally, a pre-industrial period should represent the mean climate state just before human activities started to demonstrably change the climate through combustion of fossil fuels. Here we discuss which time period might be most suitable, considering various factors such as radiative forcings, availability of observations and uncertainties in our knowledge.

We will focus on global temperatures, specifically for informing discussions on future temperature limits, and make an assessment of how much global average temperature has already warmed since our defined pre-industrial period using a range of approaches. We will also provide recommendations for: (i) how future international climate reports and agreements might use this assessment; and (ii) how the assessment itself may be improved in future, particularly regarding the use of instrumental data, proxy evidence and simulations of past climate.

[1] http://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf

Relevance of the pre-industrial period for crossing global temperature thresholds

In the absence of a formal definition for pre-industrial, the IPCC AR5 made a pragmatic choice to reference global temperature to the mean of 1850-1900 when assessing the time at which particular temperature levels would be crossed (Kirtman et al., 2013). In the final draft, 1850-1900 was referred to as 'pre-industrial', but at the IPCC AR5 plenary approval session, *'a contact group developed a proposal, in which reference to "pre-industrial" is deleted, and this was adopted [by the governments]'* (IISD, 2013). However, the term 'pre-industrial' was used in AR5, often inconsistently, in other contexts, e.g., when discussing atmospheric composition, radiative forcing (the year 1750 is used as a zero-forcing baseline), sea level rise and paleoclimate information. These discussions highlight the importance of defining pre-industrial consistently and more precisely.

In AR5, the observed increase in global temperature was calculated as the mean of 1986-2005 minus the mean of 1850-1900 in the HadCRUT4 dataset (0.61°C, Morice et al., 2012), which was the only combined global land and ocean temperature dataset available back to 1850 at the time. The 1986-2005 modern period was chosen[2] because the design of the CMIP5 simulations required a recent reference baseline for the projections of future climate (discussed further in Hawkins and Sutton, 2016). Note that the warming between 1850-1900 and the most recent decade covered (2003-2012) was given by AR5 as 0.78 ± 0.03°C (IPCC, 2013).

The choice of 1850-1900 as the historical reference period benefits from relatively widespread, but still sparse, temperature observations, and quantified uncertainties in the estimates of global temperature. Since the AR5, two further datasets have been produced that allow a comparison for the 1850-1900 period. In the Cowtan and Way (2014) dataset (hereafter CW14), which is based on interpolating the spatial gaps in HadCRUT4, the difference from 1850-1900 to 1986-

[2]The World Meteorological Organisation uses 1981-2010 for 'operational normals', which is very similar to the 1986-2005 period in terms of global mean temperature.

2005 is 0.65°C and in the Berkeley Earth global land & sea data (BEST-GL, berkeleyearth.org), it is 0.71°C, suggesting that the AR5 value may be slightly too low[3]. Also, Cowtan et al. (2015) presented GCM-based evidence that sparse observation-based datasets may have significantly underestimated the changes in global surface air temperature due to slower warming regions being preferentially sampled in the past. However, infilling the gaps in the early period is especially problematic due to the sparse observations and may accentuate the dominant observed anomaly.

However, some anthropogenic warming is estimated to have already occurred by 1850 (Hegerl et al., 2007; Schurer et al., 2013; Abrams et al., 2016) as greenhouse gas concentrations had started increasing around a century earlier (Fig. 1). On the other hand, the 1880s and 1890s were cooler than the preceding decades because of the radiative impact of aerosols from several volcanic eruptions (Fig. 1) which may have compensated for the earlier anthropogenic influence. It is therefore plausible that a 'true' pre-industrial temperature could be warmer or cooler than 1850-1900, depending on the balance of these two factors. A key question which we will consider is how representative the 1850-1900 period is for pre-industrial global average temperature.

Defining a suitable pre-industrial period using radiative forcing estimates

Anthropogenic climate change is occurring on top of: (i) internal climate variability, such as ENSO, the Pacific Decadal Oscillation (PDO), Atlantic Multi-decadal Variability (AMV) and possibly longer timescales (see Deser et al. (2010) for a review) and (ii) multi-decadal scale variations in natural radiative forcings, such as solar activity, changes in Earth's orbit and the frequency of large volcanic eruptions.

[3]These three datasets all use the Hadley Centre estimates for the sea surface temperatures since 1850 (HadSST3, Kennedy et al., 2011), and are based on similar land-based observations, so are not independent.

A pre-industrial climate should therefore be defined as a period close to present but which is before the 'industrial age', with small anthropogenic forcings. Ideally, levels of natural forcings would also be similar to present and widespread direct or indirect observations would be available. The better part of a century would appear to be required to average over the longer-timescale internal variations.

Unfortunately, such a perfect time period does not exist so compromises have to be made. In particular, there are very few instrumental temperature records before 1850 which limits our ability to determine pre-1850 global temperatures. Changes in land-use and other human activities (e.g., biomass burning, deforestation) may have altered the composition of the atmosphere several millennia ago (Ruddiman, 2003; Ruddiman et al., 2016). There are also variations in greenhouse gas concentrations (of a few ppm) before 1700 (Bauska et al., 2015). However, we assume that these early influences are not relevant for defining a pre-industrial period for use by policymakers.

Bradley et al. (2016) identified the period 725-1025 as a 'medieval quiet period', without major tropical eruptions or solar variations, and which might represent a reference climate state. However, proxy evidence suggests a slow decline of global temperatures, surface ocean temperatures and reductions in sea level over the last two millennia, which has been attributed to orbital forcing (Kaufman et al., 2009) or to increasing volcanic activity (McGregor et al., 2015; Stoffel et al., 2015; Kopp et al., 2016). Given this multi-millennial trend, whatever its cause, it makes sense to chose a reference period as close to the present as possible.

An important moment at the start of the industrial age was when James Watt patented the steam engine condenser in 1769, dramatically improving Thomas Newcomen's 1712 steam engine design. Various agricultural revolutions also began around the same time. However, there was probably only a small climate effect of these developments for several decades at least. For these reasons, historical anthropogenic radiative forcings are often considered relative to 1750 levels (Solomon et al., 2007; Meinshausen et al., 2011).

It is also important to ensure that the natural forcings in any chosen period are not unusual, compared to the present (Fig. 1). The period before 1720, often called the Little Ice Age (Mann et al., 2009), was influenced by several large tropical volcanic eruptions in the 1600s (Briffa et al., 1998; Crowley et al., 2008; Gao et al., 2008; Sigl et al., 2013) and the Maunder Minimum in solar activity which finished in the early 1700s (Steinhilber et al., 2009; Lockwood et al., 2014; Usoskin et al., 2015). The period after 1800 is influenced by the Dalton Minimum in solar activity and the large eruptions of an unlocated volcano in 1808/9, *Tambora* (1815, Raible et al., 2016), and several others in the 1820s and 1830s. In addition, greenhouse gas concentrations had already increased slightly by this time (Fig. 1).

In contrast, between 1720 and 1800 the evidence suggests that natural radiative forcings are closer to modern levels, with only very weak anthropogenic forcings. It could be argued that this period has slightly anomalously low volcanic activity, including one relatively small tropical eruption (*Makian*, Indonesia in 1761) and one long-lasting northern extra-tropical eruption (*Laki*, Iceland in 1783). This issue is returned to later.

There is also no evidence for unusual AMV/PDO variability during the 1720-1800 period (e.g., Gray et al., 2004; MacDonald and Case, 2005), suggesting that these modes of variability are not expected to significantly affect the multi-decadal temperature average.

We therefore suggest that 1720-1800 is the most suitable period to be called pre-industrial for assessing global temperature levels in terms of the radiative forcings and we concentrate on this period in the analysis which follows. Different choices may be made if considering changes in other variables (Knutti et al., 2015), such as regional temperatures, rainfall, sea level, carbon storage or glacier extents, but assessing those is beyond the scope of this study.

Using three different approaches, we now address two related questions, based on the reference periods used in IPCC AR5: (i) what is the global temperature change from our pre-industrial choice to a recent baseline (1986-2005), and (ii) is 1850-1900 a reasonable pragmatic surrogate for the pre-industrial period? We also consider the precision to which such questions can be answered.

Approach 1: using radiative forcings

Our first approach uses radiative forcings to estimate changes in global temperature before the available observations. The Coupled Model Intercomparison project, phase 5 (CMIP5) provides estimated historical radiative forcings for 1765-2005, referenced to 1750, and for a range of representative concentration pathways (RCPs) after 2005 (Meinshausen et al., 2011). We use RCP4.5 for the period 2006-2015 but this makes little difference.

We adopt a weighted least-squares multiple linear regression approach, using the radiative forcings (provided in Wm^{-2}), multiplied by individual scaling factors, to best fit the observed global mean surface temperature (GMST):

$$\mathrm{GMST}(t) = \left(\sum_{f=1}^{4} \alpha_f F_f(t) \right) + \gamma E(t-\tau) - \beta \tag{1}$$

We consider four radiative forcings (F_f, with scalings α_f): greenhouse gases, other anthropogenic effects (mainly aerosols, land use and ozone), solar, and volcanic activity. Annual means are used everywhere. We also use an ENSO index (E, scaled by γ) as a 'forcing' to remove the effects of the leading mode of interannual variability from the observations. This E index is defined as the linearly detrended Nino3.4 anomaly from 1857-2015 (Kaplan et al., 1998) and zero before 1857, with a lag (τ) of 4 months to maximise the variance explained (i.e. the annual mean is a September to August average). A similar approach to fitting global temperatures was taken by Lean and Rind

(2009) and Suckling et al. (2016). All global temperature data are referenced to 1986-2005 to match the analysis in IPCC AR5 (Kirtman et al., 2013) and β is a constant offset to account for this reference period.

We perform the analysis separately for five global temperature datasets to represent the uncertainty in temperature reconstructions, although this is an underestimate of the true uncertainty because they are all based on similar observations. For HadCRUT4, BEST-GL and CW14, the multiple linear regression is performed over the period 1850-2015. The NOAA GlobalTemp (Karl et al., 2015) and NASA GISTEMP (Hansen et al., 2010) datasets are fitted over the full extent of their available data (1880-2015). We use the HadCRUT4 uncertainties in the weighted regression (except for BEST-GL and NOAA GlobalTemp which have their own uncertainty estimates), so that the older (and more uncertain) data has less weight.

Fig. 2a shows one estimate of GMST (HadCRUT4) and the scaled forcings for the full 1765-2015 period, using the regression parameters derived over 1850-2015. The correlation between the scaled forcings (including ENSO) and observed temperatures is 0.94 for each of the global datasets.

There are two ways to estimate a change in temperature using this approach[4]. Firstly, we can average the scaled forcings over 1765-1800 to produce an estimate of the pre-industrial global temperature for each dataset with associated uncertainties, accounting for the covariance in derived α_f's. Note that this is the longest period available using the CMIP5 forcings in the 1720-1800 period. The Paleoclimate Modelling Intercomparison Project (PMIP) protocol does not currently provide consistent forcing estimates in this way for the 850-1850 period (Schmidt et al., 2012). For the five temperature datasets, the best estimates are found to range from $0.64-0.76$°C with

[4]These estimates are largely insensitive to whether a lag is introduced in the greenhouse gas forcing (as done in Lean and Rind, 2009), or if only the 1900-2015 period is used for fitting or if the anthropogenic forcings are combined before fitting.

uncertainties of around ± 0.05°C. Alternatively, the value of the regression constant (β) is an estimate of the temperature change from a state of zero forcing (in this case 1750) to 1986-2005. For the five temperature datasets, β ranges from $0.69 - 0.82$°C (with uncertainties of ± 0.02°C), which is around 0.06°C larger than using the 1765-1800 average. This difference is consistent with the small increase in greenhouse gas forcing and the relatively weak volcanic forcing after 1765. Overall, these results suggest that pre-industrial was slightly cooler than the 1850-1900 period.

Also, the derived estimates for the warming are all larger than the value used in IPCC AR5 (0.61°C), with the HadCRUT4-based estimates being the smallest and GISTEMP the largest. The differences between estimates from the various datasets are larger than the stated uncertainties, and are dominated by the uncertainty in global change since 1850, partly related to the way missing data is treated. The CW14 dataset, which interpolates between the gaps in HadCRUT4, finds slightly larger warming, consistent with Cowtan et al. (2015) who show a similar effect when examining simulated data to determine the effects of incomplete spatial coverage. The NOAA and GISTEMP datasets also use slightly different interpolation techniques. These various infilling approaches may reduce the bias from poor spatial sampling, especially for fast warming regions such as the Arctic, but may simply accentuate the dominant anomaly and add uncertainty. These inconsistencies merit further investigation elsewhere.

This approach does not account for non-linearities in the temperature response to forcings, or uncertainties in the assumed CMIP5 forcing history itself, which are likely to be particularly large for aerosols (e.g. Carslaw et al., 2013; Stevens, 2013) and ozone (Marenco et al., 1994). However, this approach does allow for varying sensitivities (α_f) to the different assumed forcings (or 'efficacies') (Hansen et al., 2005; Shindell, 2014). Another approach would be to use a simple energy balance model, tuned to the observational record (e.g. Osborn et al., 2006; Aldrin et al., 2012) and this could be examined in future work.

Approach 2: using last millennium simulations

An alternative approach to considering the forcings alone is to use 'last millennium' ensembles (LMEs) which use global climate models (GCMs) to simulate global climate from 850 to 2005 using the PMIP3 estimates of greenhouse gas concentrations, solar variations and volcanic eruptions detailed by Schmidt et al. (2012). Here we consider three ensembles with different GCMs: GISS E2-R (3 members, Schmidt et al., 2014), CESM1 (10 members, Otto-Bliesner et al., 2016) and MPI-ESM (3 members, Jungclaus et al., 2014). These are the only models to have made continuous simulations available for the whole time period using all radiative forcings[5] and multiple ensemble members (Fig. 2b).

In the GCM simulations, 1720-1800 is $0.00-0.06$°C cooler than 1850-1900 (using ensemble means), which is slightly smaller than the result using Approach 1. However, the three GCMs produce very different estimates for the warming from 1720-1800 until 1986-2005 (0.51 ± 0.08°C for CESM1, 1.04 ± 0.07°C for GISS E2-R and 0.91 ± 0.04°C for MPI-ESM)[6]. These differences are not what would be expected due to climate sensitivity alone as CESM1 has the largest transient climate response (TCR, 2.2K) and GISS E2-R the smallest (1.5K). It is more likely that the differences are due to a combination of several factors, including climate sensitivity, different amplitude responses to anthropogenic aerosols and volcanic eruptions (Stoffel et al., 2015), different assumed forcings (e.g., the size of the 1761 eruption), and different implementations of the forcings. In addition, the global temperature response to volcanic eruptions appears to be larger in the GCMs than the real world (e.g. Schurer et al., 2013), although Stoffel et al. (2015) suggest this effect is much reduced with an improved representation of the aerosol microphysics.

---

[5]Note that the GISS E2-R simulations used a different aerosol forcing over the historical period than the CMIP5 historical simulations performed with the same GCM. The PMIP3 simulations warm by about 0.3K more than the CMIP5 simulations (not shown).

[6]We also tested Approach 1 using the global temperatures from the PMIP simulations. This produced compatible values for the warming (0.45 ± 0.09, 1.09 ± 0.04° and 0.90 ± 0.06°C respectively), building confidence in that approach.

Given the diversity in global temperature response, a robust estimate of change in global temperature since pre-industrial using these simulations should consider scaling the responses to the observations or using detection and attribution techniques on the range of simulations available (Schurer et al., 2013; Otto-Bliesner et al., 2016). In addition, the comparison with observations is not necessarily like-with-like given sparse observations and different use of air or sea temperatures (Cowtan et al., 2015; Richardson et al., 2016).

However, an additional use for the LMEs is to examine uncertainty in the estimate of pre-industrial temperatures due to internal variability alone. This can be done by considering the spread of estimated change using the ten CESM1 ensemble members ($\sigma = 0.05$K), which suggests an uncertainty of around $\pm 0.1°$C. Note that this range is similar to the uncertainty ranges from long instrumental records discussed below. The other ensembles are too small to reliably estimate this range. We also use the CESM1 simulations to consider issues of differential seasonal warming in the Appendix.

Approach 3: using long instrumental records

The above two approaches have considered the response to estimated radiative forcings. An alternative approach to estimate GMST further back in time is to use direct observations from long instrumental records and calibrate them against each of the five global mean temperature datasets.

For example, Central England Temperature (HadCET, Manley, 1974; Parker et al., 1992, hereafter referred to as CET) is available for 1659-present. CET covers just 0.005% of the Earth's surface but is highly correlated with GMST on multi-decadal timescales (Sutton et al., 2015). Here, we utilise this correlation and scale GMST to CET:

$$\mathrm{CET} = \delta \mathrm{GMST} + \varepsilon \tag{2}$$

using the overlapping periods (1850-2015), and adopt the same parameters to scale CET back to 1659 as an estimate of GMST (Fig. 3a). When using HadCRUT4 as GMST, $\delta = 1.20 \pm 0.23$, although other global temperature datasets give lower values (e.g., for BEST-GL, $\delta = 1.06 \pm 0.21$). The major caveats to this approach are that we assume the historical temperature estimates are unbiased, and that the relationship between GMST and CET is the same whatever forcing is dominant, neither of which may be true (Zanchettin et al., 2013; Haarsma et al., 2013, and see Appendix).

We take the mean of the scaled CET over two periods: (i) 1765-1800 (for consistency with Approach 1) and (ii) 1720-1800 (the full period identified from the radiative forcing history). An additional issue that arises from scaling a local record to global temperatures is the possible regional effect of external forcing. In particular, the eruption of *Laki* (located in Iceland) in 1783 likely only had a small global effect, but it certainly influenced western Europe (Thordarson and Self, 2003). Therefore the years 1783 and 1784 are removed from the averages due to the eruption of *Laki* to avoid biasing the estimated temperature change. However, this does not change the results significantly.

These two periods produce consistent estimates for the warming to 1986-2005: $0.75 \pm 0.10\,^{\circ}$C (for 1765-1800) and $0.64 \pm 0.08\,^{\circ}$C (for 1720-1800) when using HadCRUT4 for GMST. The other global temperature datasets give larger values for the warming to 1986-2005, by up to 0.09°C (Fig. 3a). The quoted uncertainty ranges account for the uncertainties in the regression parameters and assume the uncertainty in each CET annual mean from 1720-1800 is independent and equal to 0.2°C (based on Parker, 2010).

The difference between the two averaging periods is mainly because the 1720s and 1730s were unusually warm in the CET record. Internal climate variability and a recovery from the negative forcings of the previous decades are possible explanations, although this warmth was less pronounced in some other European instrumental records (e.g. Berlin) (Jones and Briffa, 2006).

Figs. 3b repeats this analysis with the Berkeley global land temperature (BEST-Land, Rohde et al., 2013), which starts in 1753. A similar approach was adopted by Mann (2014). Using BEST-Land produces a consistent but slightly lower warming than derived with CET. Using the scaled temperatures over the 1753-1800 period, the estimates of the warming to 1986-2005 range from 0.62 ± 0.10°C for HadCRUT4 to 0.71 ± 0.12°C for GISTEMP.

It may seem surprising that the error bars are not smaller for the BEST-Land dataset than for CET. The regression uncertainty is indeed much larger for the local example, however the error in representing the whole global land area with sparse data is larger than in representing central England with a small number of stations. These two sources of uncertainty combine to give similar overall ranges. Note that BEST-Land looks very similar to the long European records and the variability increases further back in time (also for CET), highlighting that fewer and fewer (mostly European) stations are used in the reconstruction.

We also consider a long temperature series from the Netherlands, referenced to De Bilt, which starts in 1706 (Van Engelen and Nellestijn, 1990) and a Central Europe instrumental series from Dobrovolnỳ et al. (2010) which starts in 1760, which are also both well correlated with GMST in the overlapping period. These results are summarised in Fig. 4 which shows that the Central Europe series consistently produces slightly lower estimates of the warming than CET or BEST-Land.

Overall assessment

We consider that approaches based on the radiative forcings and scaled instrumental observations currently produce more reliable estimates of the global temperature change since pre-industrial than the last millennium GCM simulations. This weighting of methods could change in future with additional evidence, analysis and model development (see implications discussed below). Furthermore, the estimates using radiative forcings are generally larger than when using the observational datasets, as summarised in Fig. 4. Much of the uncertainty in the assessment derives from the range of global temperature change estimates available since 1850. For example, the uninterpolated HadCRUT4 dataset produces lower values than the other infilled records.

Our overall assessment is that the change in global average temperature from pre-industrial to 1986-2005 is 'likely' between $0.55 - 0.80\,^{\circ}$C.

This range reflects the authors' aggregated assessment of the three approaches and contains virtually all of the best estimates using the various combinations of regional and global temperature datasets and scaled radiative forcing estimates. Note that there are potentially important uncertainties in each approach which we cannot quantify. As in IPCC AR5 we consider that 'likely' refers to greater than 66% probability, although this is not a formal uncertainty quantification.

It is also helpful to assess a lower bound and we suggest that the warming since pre-industrial is 'likely' greater than $0.60\,^{\circ}$C, implying that the value used by IPCC AR5 for the warming since 1850-1900 (0.61°C) was probably smaller than the true change since pre-industrial. Such differences matter more when considering the chance of crossing lower temperature levels such as 1.5°C than when considering higher values.

Using this lower bound, 2015 was the first year to be more than 1°C above pre-industrial levels in each global temperature dataset (Fig. 5). 2016 is currently on track to be warmer than 2015, but future years could still be cooler than 2015 due to internal variability, such as a La Niña event.

The available proxy-based evidence is consistent with our assessment, but currently too uncertain to make more precise estimates, partly due to different seasonal signals (see Appendix).

However, defining a pre-industrial period offers a target for proxy reconstructions to aid future assessments.

Conclusions & implications

We have examined estimates of historical radiative forcings to determine which period might be most suitable to be termed pre-industrial and used several approaches to estimate a change in global temperature since this pre-industrial reference period. The main conclusions are:

1. The 1720-1800 period is most suitable to be defined as pre-industrial in physical terms, although we have incomplete information about the radiative forcings and very few direct observations during this time. However, this definition offers a target period for future analysis and data collection to inform this issue.

2. The 1850-1900 period is a reasonable pragmatic surrogate for pre-industrial global mean temperature. The available evidence suggests it was slightly warmer than 1720-1800 by around 0.05°C, but this is not statistically significant.

3. We assess a 'likely' range of 0.55 – 0.80 °C for the change in global average temperature from pre-industrial to 1986-2005.

4. We also consider a likely lower bound on warming from pre-industrial to 1986-2005 of 0.60°C, implying that the AR5 estimate of warming was probably too small and that 2015 was the first year to be more than 1°C above pre-industrial levels.

We have assumed in the motivation for this discussion and choice of reference periods that the UNFCCC agreements on temperature limits refer to anthropogenic increases only, but this is not explicitly stated. We have not attempted to attribute the observed increase in global temperatures (but see Schurer et al., 2013; Otto et al., 2015); non-anthropogenic factors (including internal variability) may have either offset or contributed to the warming. We have attempted to minimise issues of varying natural forcing and internal variability, but this effect cannot be removed entirely.

Our chosen pre-industrial period likely has slightly weaker volcanic activity than a typical period and the modern reference period (1986-2005) includes the large Pinatubo eruption. These effects would bias our estimated change in temperature to be slightly too low, highlighting the value of assessing a lower bound in the warming since pre-industrial. We also note that future climate projections do not usually include volcanic eruptions so choosing a relatively weak volcanic baseline is perhaps appropriate. The recent period has a slightly positive PDO index which would act as a small positive bias for some of our estimates, but this modern reference period will likely be updated for the next IPCC assessment.

There are a number of ways that this assessment could be improved. Better understanding of historical radiative forcings, particularly of volcanic eruptions, solar activity and anthropogenic aerosols, would help narrow the uncertainties in past global and regional temperature change. We did not include the estimates for pre-industrial temperature from the last millennium simulations in this assessment due to the diverse derived values, which is due to differences in both the forcings used and climate sensitivity (Fernández-Donado et al., 2013). Future work might consider scaling the simulations (Schurer et al., 2013) or use of simple Energy Balance Models (EBMs).

However, we may not necessarily expect simulated and observed values to agree, even in the case of perfect knowledge of radiative forcings and climate sensitivity. This is because the global observations are a sparse blend of sea surface temperatures over the ocean and air temperatures over the land, whereas virtually all analyses of GCM simulations use air temperatures with complete global coverage. Cowtan et al. (2015) and Richardson et al. (2016) used GCM simulations to suggest that if we had complete coverage of air temperature, the observed change from 1850 to present would be $24 \pm 15\%$ larger than currently estimated in HadCRUT4. The use of infilled temperature datasets only partly overcomes this issue.

This creates a dilemma - are the temperature limits adopted by the UNFCCC designed to use observationally-based estimates of global temperature change (as generally used here) or on what those observations mean for a 'true' global mean air temperature change (as used in most climate impact assessments)? The available evidence suggests that the latter is larger. If such findings are borne out by further research, and if the 'true' change is what is desired by UNFCCC, then our assessed temperature change since pre-industrial is too small and should probably be increased by $0.05-0.10°$C.

It is possible to obtain additional data for the historical period. Recovery of additional instrumental observations of temperature and sea level pressure from undigitised hand-written logbooks from ships and in currently data sparse regions could significantly aid similar future assessments. Some such efforts are ongoing (e.g. the ACRE and OldWeather.org initiatives, Allan et al., 2011) but these could be expanded. The available observations can also be combined with data assimilation techniques to allow longer atmospheric reanalyses to be produced (Widmann et al., 2010; Compo et al., 2011; Matsikaris et al., 2015; Brohan et al., 2016). Additional seasonal proxy information would be of great value for informing this discussion, especially for winter (see Appendix) and for the tropics and Southern Hemisphere (e.g. Jones et al., 2016), although the temporal resolution and continuity of proxies into the modern period is also a potential issue. Also note that a suitable pre-industrial period may be different for other climate variables, e.g. sea level, or for carbon cycle considerations.

Two specific recommendations for future GCM-based analyses and simulations are: (i) to use blended observation-like estimates of global mean temperature when comparing observations and simulations, and (ii) use 1750 forcings to perform pre-industrial control simulations and to start historical transient simulations, rather than 1850. Adopting these recommendations would allow an ensemble of transient historical simulations to better quantify the role of natural variability and the impacts of the total radiative forcing changes since the pre-industrial period, especially the potentially long-term impact of the large volcanic eruptions in the early 1800s (Raible et al., 2016). We recognise, however, that this increases the computational demand in producing historical simulations. In addition, increased usage of tracers (e.g. water stable isotopes) and proxy models within GCMs would allow more direct comparisons between simulations and proxy observations, including GCM simulations nudged to atmospheric reanalyses (e.g. Jouzel et al., 2000; LeGrande and Schmidt, 2009; Evans et al., 2013).

Finally, these findings have a number of implications for policy-relevant issues. For example, the date at which future temperature thresholds are expected to be crossed may be shifted slightly earlier than estimated in IPCC AR5 (see Joshi et al., 2011; Kirtman et al., 2013; Hawkins and Sutton, 2016). In addition, the cumulative emissions allowed to avoid reaching a particular temperature threshold (Meinshausen et al., 2009; Allen et al., 2009) may need to be reassessed, although any difference would likely be well within the current uncertainty ranges. Moving the baseline may also affect how historical responsibility for emissions needs to be accounted for (Knutti et al., 2015).

More specifically, given the uncertainty in the global mean temperature change since pre-industrial, the UNFCCC might consider alternative equivalent baselines and limits to global temperature change. For example, *"well below* 2°C *above pre-industrial"* might be translated to *"well below* X°C *above 1986-2005"*. Using a recent baseline is possibly more relevant for defining some impacts of climatic changes, with the value of X (and choice of baseline period) being decided by the UNFCCC. Given the uncertainty in defining the temperature change since pre-industrial, such a framing would allow a more precise assessment of when such levels might be reached in future, given our much improved recent observational coverage and availability of atmospheric reanalyses for the modern period (e.g. Dee et al., 2011; Simmons et al., 2016). It would also remove the need to precisely assess inherently uncertain changes since the pre-industrial period.

APPENDIX

Comparison with proxy reconstructions

There are numerous efforts to reconstruct past climate using different proxies and archives which could be used to aid an assessment of change since the pre-industrial period. For temperature, these include ice cores, glaciers, tree rings, pollen, corals and sediment cores.

For example, Leclercq and Oerlemans (2012) suggest a global land warming of 0.94 ± 0.31°C between 1830 and 2000 using glacier reconstructions, although the mid-1700s is around 0.25°C warmer than 1830 in their estimates. Pollack and Smerdon (2004) suggest that global land temperatures in the mid-1700s were around $0.65-0.90$°C below the year 2000 using borehole proxies. Mann et al. (2008) perform a multi-proxy analysis and report that global average temperature was around 0.3°C below 1961-90 in the mid-1700s, with large uncertainties. This is equivalent to around 0.6°C below 1986-2005, consistent with the recent PAGES2k global reconstruction (PAGES 2k Consortium et al., 2013) and this study.

Overall, these proxy reconstruction estimates for pre-industrial temperature are consistent with the approaches adopted above, but the uncertainties are currently too large to make more precise statements. Defining a pre-industrial period (1720-1800) will hopefully provide a target for future reconstructions using the proxy data available. Certain long proxy series could also be used in Approach 3. However, it is important that such efforts focus on all seasons, as we next discuss.

Seasonal effects in proxies, observations and simulations

There are likely some seasonal differences in the rates of temperature change which are important to consider (e.g. Hegerl et al., 2011; Jones et al., 2014). For example, different proxies are sensitive to climate in certain seasons. In general, summer is more widely represented because many proxies rely on biological activity which tends to occur in the extended summer season. This is a potential issue for using proxies to reconstruct past temperatures if winter and summer change at different rates (Jones et al., 2003). In that case, the different seasonal proxies may not agree and/or produce biased estimates of an annual average. Some reconstructions (e.g. Van Engelen et al., 2001; Luterbacher et al., 2004; Vinther et al., 2010) for Holland, Europe and Greenland respectively do show seasonal warming differences. However, the restricted availability of winter proxies limits the scope of such a comparison.

To investigate how representative of annual mean changes the seasonal data is, we repeated the instrumental analysis (Approach 3) using extended seasons (April to September and October to March) for the regional data, whilst retaining the annual global data as the reference. Fig. 6a shows how the derived warming since the 1753-1800 period depends on the choice of season for the instrumental series - the extended winter season warms much faster than the extended summer season.

However, if this seasonal difference in the rate of change over Europe was constant with time it should be scaled out. This suggests that there is: (i) a seasonal bias in the observed temperatures in certain periods (e.g. before standardised measurements) and/or (ii) a different seasonal response to different radiative forcings.

For example, there is evidence that some historical observations may be biased, especially in summer, where warm biases due to non-optimal observation techniques in the past have been identified (Parker, 1994; Böhm et al., 2010; Jones, 2016), which fits the pattern seen in Fig. 6a. Dobrovolnỳ et al. (2010) note that their documentary temperature data agrees best with their instrumental data during winter, adding credence to this hypothesis. In addition, the cooling due to tropospheric aerosols in the 20th century may be seasonally dependent (Hunter et al., 1993; Krishnan and Ramanathan, 2002), there is a trend in westerly wind characteristics in winter (Haarsma et al., 2013) and many of the observations are located in the northern extra-tropics and therefore influenced by Arctic amplification, which is observed and simulated to be larger in winter than in summer (Serreze et al., 2009; Pithan and Mauritsen, 2014).

We can also examine whether this seasonal warming difference is present in the last millennium model simulations. Fig. 6b highlights that the CESM1 LME simulations do not show a strong global mean warming seasonal difference since the pre-industrial period, and only a very small seasonal effect when considering the central England location. The complex nature of these different seasonal features merits further analysis in a range of observations and simulations.

Acknowledgments. We thank John Fasullo and Johann Jungclaus for providing the CESM1 LME and MPI-ESM data respectively. EH is funded by the UK National Centre for Atmospheric Science (NCAS) and a Natural Environment Research Council (NERC) fellowship (Grant NE/I020792/1). TJO and GH were supported by the NERC (grant number NE/N006348/1, SMURPHS). ES was supported by the European Union Seventh Framework Programme (FP7/2007-2013) under the SPECS project (grant Agreement No. 308378) and by the UK-China Research and Innovation Partnership Fund through the Met Office Climate Science for Service Partnership (CSSP) China as part of the Newton Fund. AS and GH are supported by the ERC funded project TITAN (EC-320691). GH was further supported by NCAS and the

Wolfson Foundation and the Royal Society as a Royal Society Wolfson Research Merit Award (WM130060) holder.

References

Abrams, N., H. McGregor, J. Tierney, M. Evans, N. McKay, D. Kaufman, and the PAGES consortium, 2016: Early onset of industrial-era warming across the oceans and continents. *Nature*, **536**, 411–418.

Aldrin, M., M. Holden, P. Guttorp, R. B. Skeie, G. Myhre, and T. K. Berntsen, 2012: Bayesian estimation of climate sensitivity based on a simple climate model fitted to observations of hemispheric temperatures and global ocean heat content. *Environmetrics*, **23 (3)**, 253–271, doi: 10.1002/env.2140.

Allan, R., P. Brohan, G. P. Compo, R. Stone, J. Luterbacher, and S. Bronnimann, 2011: The international atmospheric circulation reconstructions over the earth (ACRE) initiative. *Bull. Amer. Meteor. Soc.*, **92**, 1421, doi:10.1175/2011BAMS3218.1.

Allen, M. R., D. J. Frame, C. Huntingford, C. D. Jones, J. A. Lowe, M. Meinshausen, and N. Meinshausen, 2009: Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature*, **458**, 1163–1166.

Bauska, T. K., F. Joos, A. C. Mix, R. Roth, J. Ahn, and E. J. Brook, 2015: Links between atmospheric carbon dioxide, the land carbon reservoir and climate over the past millennium. *Nature Geoscience*, **8**, 383–387, doi:10.1038/ngeo2422.

Böhm, R., P. D. Jones, J. Hiebl, D. Frank, M. Brunetti, and M. Maugeri, 2010: The early instrumental warm-bias: a solution for long central european temperature series 1760–2007. *Climatic Change*, **101**, 41–67.

Bradley, R. S., H. Wanner, and H. F. Diaz, 2016: The medieval quiet period. *The Holocene*, in press, doi:10.1177/0959683615622552.

Briffa, K. R., P. D. Jones, F. H. Schweingruber, and T. J. Osborn, 1998: Influence of volcanic eruptions on northern hemisphere summer temperature over the past 600 years. *Nature*, **393**, 450–455, doi:10.1038/30943.

Brohan, P., G. P. Compo, S. Brönnimann, R. J. Allan, R. Auchmann, Y. Brugnara, P. D. Sardeshmukh, and J. S. Whitaker, 2016: The 1816 year without a summer in an atmospheric reanalysis. *Climate of the Past Discussions*, **2016**, 1–11, doi:10.5194/cp-2016-78.

Carslaw, K. S., and Coauthors, 2013: Large contribution of natural aerosols to uncertainty in indirect forcing. *Nature*, **503**, 67–71, doi:10.1038/nature12674.

Compo, G. P., and Coauthors, 2011: The twentieth century reanalysis project. *Quarterly Journal of the Royal Meteorological Society*, **137**, 1–28.

Cowtan, K., and R. G. Way, 2014: Coverage bias in the hadcrut4 temperature series and its impact on recent temperature trends. *QJRMS*, **140**, 1935–1944, doi:10.1002/qj.2297.

Cowtan, K., and Coauthors, 2015: Robust comparison of climate models with observations using blended land air and ocean sea surface temperatures. *Geophysical Research Letters*, **42 (15)**, 6526–6534.

Crowley, T., and M. Unterman, 2013: Technical details concerning development of a 1200 yr proxy index for global volcanism. *Earth System Science Data*, **5 (1)**, 187–197.

Crowley, T. J., G. Zielinski, B. Vinther, R. Udisti, K. Kreutz, J. Cole-Dai, and E. Castellano, 2008: Volcanism and the little ice age. *PAGES news*, **16 (2)**, 22–23.

Dee, D. P., S. M. Uppala, A. J. Simmons, and et al., 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *QJRMS*, **137**, 553–597, doi: 10.1002/qj.828.

Deser, C., M. A. Alexander, S.-P. Xie, and A. S. Phillips, 2010: Sea surface temperature variability: Patterns and mechanisms. *Annual Review of Marine Science*, **2 (1)**, 115–143, doi: 10.1146/annurev-marine-120408-151453.

Dobrovolnỳ, P., and Coauthors, 2010: Monthly, seasonal and annual temperature reconstructions for central europe derived from documentary evidence and instrumental records since ad 1500. *Climatic Change*, **101 (1-2)**, 69–107.

Evans, M., S. Tolwinski-Ward, D. Thompson, and K. Anchukaitis, 2013: Applications of proxy system modeling in high resolution paleoclimatology. *Quaternary Science Reviews*, **76**, 16 – 28, doi:http://dx.doi.org/10.1016/j.quascirev.2013.05.024.

Fernández-Donado, L., and Coauthors, 2013: Large-scale temperature response to external forcing in simulations and reconstructions of the last millennium. *Climate of the Past*, **9 (1)**, 393–421, doi:10.5194/cp-9-393-2013.

Gao, C., A. Robock, and C. Ammann, 2008: Volcanic forcing of climate over the past 1500 years: An improved ice core-based index for climate models. *Journal of Geophysical Research: Atmospheres*, **113**, doi:10.1029/2008JD010239, d23111.

Gray, S. T., L. J. Graumlich, J. L. Betancourt, and G. T. Pederson, 2004: A tree-ring based reconstruction of the atlantic multidecadal oscillation since 1567 ad. *Geophysical Research Letters*, **31 (12)**.

Haarsma, R. J., F. Selten, and G. J. van Oldenborgh, 2013: Anthropogenic changes of the thermal and zonal flow structure over Western Europe and Eastern North Atlantic in CMIP3 and CMIP5 models. *Clim. Dyn.*, **41 (9-10)**, 2577–2588, doi:10.1007/s00382-013-1734-8.

Hansen, J., R. Ruedy, M. Sato, and K. Lo, 2010: Global surface temperature change. *Reviews of Geophysics*, **48 (4)**.

Hansen, J., and Coauthors, 2005: Efficacy of climate forcings. *Journal of Geophysical Research: Atmospheres*, **110**, D18104, doi:10.1029/2005JD005776.

Hawkins, E., and R. Sutton, 2016: Connecting climate model projections of global temperature change with the real world. *Bulletin of the American Meteorological Society*, **97 (6)**, 963–980, doi:10.1175/BAMS-D-14-00154.1.

Hegerl, G., J. Luterbacher, F. Gonzlez-Rouco, S. F. B. Tett, T. Crowley, and E. Xoplaki, 2011: Influence of human and natural forcing on european seasonal temperatures. *Nature Geoscience*, **4**, 99103, doi:10.1038/ngeo1057.

Hegerl, G. C., T. J. Crowley, M. Allen, W. T. Hyde, H. N. Pollack, J. Smerdon, and E. Zorita, 2007: Detection of human influence on a new, validated 1500-year temperature reconstruction. *Journal of Climate*, **20 (4)**, 650–666, doi:10.1175/JCLI4011.1.

Hunter, D. E., S. E. Schwartz, R. Wagener, and C. M. Benkovitz, 1993: Seasonal, latitudinal, and secular variations in temperature trend: Evidence for influence of anthropogenic sulfate. *Geophysical Research Letters*, **20 (22)**, 2455–2458, doi:10.1029/93GL02808.

IISD, 2013: Earth negotiations bulletin: Summary of the 12th session of working group i of the intergovernmental panel on climate change (ipcc) and thirty-sixth session of the ipcc, URL http://www.iisd.ca/vol12/enb12581e.html.

IPCC, 2013: Summary for policymakers. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley, Eds., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, book section SPM, 12̌01330, doi:10.1017/CBO9781107415324.004, URL www.climatechange2013.org.

Jones, J., S. Gille, H. Goosse, N. Abram, P. Canziani, D. Charman, and co-authors, 2016: Assessing recent trends in high-latitude southern hemisphere surface climate. *Nature Climate Change*, in press.

Jones, P., 2016: The reliability of global and hemispheric surface temperature records. *Advances in Atmospheric Sciences*, **33 (3)**, 269–282.

Jones, P., and K. Briffa, 2006: Unusual climate in northwest europe during the period 1730 to 1745 based on instrumental and documentary data. *Climatic Change*, **79 (3-4)**, 361–379.

Jones, P. D., K. R. Briffa, and T. J. Osborn, 2003: Changes in the northern hemisphere annual cycle: Implications for paleoclimatology? *Journal of Geophysical Research: Atmospheres*, **108**, 4588, doi:10.1029/2003JD003695.

Jones, P. D., C. Harpham, and B. M. Vinther, 2014: Winter-responding proxy temperature reconstructions and the north atlantic oscillation. *Journal of Geophysical Research: Atmospheres*, **119 (11)**, 6497–6505, doi:10.1002/2014JD021561.

Joshi, M., E. Hawkins, R. Sutton, J. Lowe, and D. Frame, 2011: Projections of when temperature change will exceed 2°C above pre-industrial levels. *Nature Climate Change*, **1**, 407–412, doi:10.1038/nclimate1261.

Jouzel, J., G. Hoffmann, R. Koster, and V. Masson, 2000: Water isotopes in precipitation: data/model comparison for present-day and past climates. *Quaternary Science Reviews*, **19 (1)**, 363–379.

Jungclaus, J. H., K. Lohmann, and D. Zanchettin, 2014: Enhanced 20th-century heat transfer to the arctic simulated in the context of climate variations over the last millennium. *Climate of the Past*, **10 (6)**, 2201–2213, doi:10.5194/cp-10-2201-2014.

Kaplan, A., M. Cane, Y. Kushnir, A. Clement, M. Blumenthal, and B. Rajagopalan, 1998: Analyses of global sea surface temperature 1856-1991. *Journal of Geophysical Research*, **103**, 18 567–18 589.

Karl, T. R., and Coauthors, 2015: Possible artifacts of data biases in the recent global surface warming hiatus. *Science*, **348 (6242)**, 1469–1472.

Kaufman, D. S., and Coauthors, 2009: Recent warming reverses long-term arctic cooling. *Science*, **325 (5945)**, 1236–1239, doi:10.1126/science.1173983.

Keeling, C. D., S. C. Piper, R. B. Bacastow, M. Wahlen, T. P. Whorf, M. Heimann, and H. A. Meijer, 2001: *Exchanges of atmospheric CO2 and 13CO2 with the terrestrial biosphere and oceans from 1978 to 2000. I. Global aspects*. SIO Reference Series, No. 01-06, Scripps Institution of Oceanography, San Diego.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby, 2011: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. measurement and sampling uncertainties. *Journal of Geophysical Research: Atmospheres*, **116**, doi:10.1029/2010JD015218, d14103.

Kirtman, B., and Coauthors, 2013: Near-term climate change: Projections and predictability. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley, Eds., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, chapter 11, 953–1028, doi:10.1017/CBO9781107415324.023.

Knutti, R., J. Rogelj, J. Sedlacek, and E. M. Fischer, 2015: A scientific critique of the two-degree climate change target. *Nature Geoscience*, **9**, 13–18, doi:10.1038/ngeo2595.

Kopp, R. E., and Coauthors, 2016: Temperature-driven global sea-level variability in the common era. *Proceedings of the National Academy of Sciences*, in press, doi:10.1073/pnas.1517056113.

Krishnan, R., and V. Ramanathan, 2002: Evidence of surface cooling from absorbing aerosols. *Geophysical Research Letters*, **29 (9)**, 541–544, doi:10.1029/2002GL014687.

Lean, J. L., and D. H. Rind, 2009: How will Earth's surface temperature change in future decades? *Geophys. Res. Lett.*, **36**, L15 708, doi:10.1029/2009GL038932.

Leclercq, P. W., and J. Oerlemans, 2012: Global and hemispheric temperature reconstruction from glacier length fluctuations. *Climate Dynamics*, **38 (5-6)**, 1065–1079.

LeGrande, A., and G. Schmidt, 2009: Sources of holocene variability of oxygen isotopes in paleoclimate archives. *Climate of the Past*, **5 (3)**, 441–455.

Lockwood, M., M. J. Owens, and L. Barnard, 2014: Centennial variations in sunspot number, open solar flux, and streamer belt width: 2. comparison with the geomagnetic data. *Journal of Geophysical Research: Space Physics*, **119 (7)**, 5183–5192, doi:10.1002/2014JA019972, 2014JA019972.

Luterbacher, J., D. Dietrich, E. Xoplaki, M. Grosjean, and H. Wanner, 2004: European seasonal and annual temperature variability, trends, and extremes since 1500. *Science*, **303 (5663)**, 1499–1503.

MacDonald, G. M., and R. A. Case, 2005: Variations in the pacific decadal oscillation over the past millennium. *Geophysical Research Letters*, **32 (8)**.

MacFarling Meure, C., D. Etheridge, C. Trudinger, P. Steele, R. Langenfelds, T. Van Ommen, A. Smith, and J. Elkins, 2006: Law dome co2, ch4 and n2o ice core records extended to 2000 years bp. *Geophysical Research Letters*, **33 (14)**.

Manley, G., 1974: Central england temperatures: monthly means 1659 to 1973. *Quarterly Journal of the Royal Meteorological Society*, **100 (425)**, 389–405.

Mann, M., 2014: Earth will cross the climate danger threshold by 2036. *Scientific American*, http://www.scientificamerican.com/article/earth-will-cross-the-climate-danger-threshold-by-2036/.

Mann, M. E., Z. Zhang, M. K. Hughes, R. S. Bradley, S. K. Miller, S. Rutherford, and F. Ni, 2008: Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proceedings of the National Academy of Sciences*, **105 (36)**, 13 252–13 257.

Mann, M. E., and Coauthors, 2009: Global signatures and dynamical origins of the little ice age and medieval climate anomaly. *Science*, **326**, 1256–1260, doi:10.1126/science.1177303.

Marenco, A., H. Gouget, P. Ndlec, J.-P. Pags, and F. Karcher, 1994: Evidence of a long-term increase in tropospheric ozone from pic du midi data series: Consequences: Positive radiative forcing. *Journal of Geophysical Research: Atmospheres*, **99**, 16 617–16 632, doi:10.1029/94JD00021.

Matsikaris, A., M. Widmann, and J. Jungclaus, 2015: Assimilating continental mean temperatures to reconstruct the climate of the late pre-industrial period. *Climate Dynamics*, in press, doi: 10.1007/s00382-015-2785-9.

McGregor, H. V., and Coauthors, 2015: Robust global ocean cooling trend for the pre-industrial common era. *Nature Geoscience*, **8**, 671 – 677, doi:10.1038/ngeo2510.

Meinshausen, M., N. Meinshausen, W. Hare, S. C. Raper, K. Frieler, R. Knutti, D. J. Frame, and M. R. Allen, 2009: Greenhouse-gas emission targets for limiting global warming to 2 c. *Nature*, **458 (7242)**, 1158–1162.

Meinshausen, M., and Coauthors, 2011: The rcp greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change*, **109 (1-2)**, 213–241.

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones, 2012: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *J. Geophy. Res. Atmos.*, **117**, D08 101, doi:10.1029/2011JD017187.

Osborn, T. J., S. C. B. Raper, and K. R. Briffa, 2006: Simulated climate change during the last 1,000years: comparing the echo-g general circulation model with the magicc simple climate model. *Climate Dynamics*, **27 (2)**, 185–197, doi:10.1007/s00382-006-0129-5.

Otto, F. E., D. J. Frame, A. Otto, and M. R. Allen, 2015: Embracing uncertainty in climate change policy. *Nature Climate Change*, **5**, 917920, doi:10.1038/nclimate2716.

Otto-Bliesner, B. L., and Coauthors, 2016: Climate variability and change since 850 ce: An ensemble approach with the community earth system model (cesm). *Bulletin of the American Meteorological Society*, in press.

PAGES 2k Consortium, and Coauthors, 2013: Continental-scale temperature variability during the past two millennia. *Nature Geoscience*, **6 (5)**, 339, doi:10.1038/ngeo1797.

Parker, D., 1994: Effects of changing exposure of thermometers at land stations. *International Journal of Climatology*, **14 (1)**, 1–31.

Parker, D. E., 2010: Uncertainties in early central england temperatures. *International Journal of Climatology*, **30 (8)**, 1105–1113, doi:10.1002/joc.1967.

Parker, D. E., T. P. Legg, and C. K. Folland, 1992: A new daily central england temperature series, 1772–1991. *International Journal of Climatology*, **12 (4)**, 317–342.

Pithan, F., and T. Mauritsen, 2014: Arctic amplification dominated by temperature feedbacks in contemporary climate models. *Nature Geoscience*, **7**, 181–184, doi:10.1038/ngeo2071.

Pollack, H. N., and J. E. Smerdon, 2004: Borehole climate reconstructions: Spatial structure and hemispheric averages. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **109 (D11)**.

Raible, C. C., and Coauthors, 2016: Tambora 1815 as a test case for high impact volcanic eruptions: Earth system effects. *Wiley Interdisciplinary Reviews: Climate Change*, in press, doi: 10.1002/wcc.407.

Richardson, M., K. Cowtan, E. Hawkins, and M. Stolpe, 2016: Reconciled climate response estimates from climate models and the energy budget of earth. *Nature Climate Change*, in press, doi:10.1038/nclimate3066.

Rohde, R., and Coauthors, 2013: A new estimate of the average earth surface land temperature spanning 1753 to 2011. *Geoinfor. Geostat.*, **1**, doi:10.4172/2327-4581.1000101.

Ruddiman, W. F., 2003: The anthropogenic greenhouse era began thousands of years ago. *Climatic Change*, **61 (3)**, 261–293, doi:10.1023/B:CLIM.0000004577.17928.fa.

Ruddiman, W. F., and Coauthors, 2016: Late holocene climate: Natural or anthropogenic? *Reviews of Geophysics*, **54 (1)**, 93–118, doi:10.1002/2015RG000503.

Schmidt, G., and Coauthors, 2012: Climate forcing reconstructions for use in pmip simulations of the last millennium (v1. 1). *Geoscientific Model Development*, **5**, 185–191.

Schmidt, G. A., and Coauthors, 2014: Configuration and assessment of the giss modele2 contributions to the cmip5 archive. *Journal of Advances in Modeling Earth Systems*, **6**, 141–184, doi:10.1002/2013MS000265.

Schurer, A. P., G. C. Hegerl, M. E. Mann, S. F. B. Tett, and S. J. Phipps, 2013: Separating forced from chaotic climate variability over the past millennium. *Journal of Climate*, **26**, 6954–6973, doi:10.1175/JCLI-D-12-00826.1.

Serreze, M., A. Barrett, J. Stroeve, D. Kindig, and M. Holland, 2009: The emergence of surface-based arctic amplification. *The Cryosphere*, **3 (1)**, 11–19.

Shindell, D. T., 2014: Inhomogeneous forcing and transient climate sensitivity. *Nature Climate Change*, **4 (4)**, 274–277.

Sigl, M., and Coauthors, 2013: A new bipolar ice core record of volcanism from wais divide and neem and implications for climate forcing of the last 2000years. *Journal of Geophysical Research: Atmospheres*, **118 (3)**, 1151–1169, doi:10.1029/2012JD018603.

Simmons, A. J., P. Berrisford, D. P. Dee, H. Hersbach, S. Hirahara, and J.-N. Thpaut, 2016: A reassessment of temperature variations and trends from global reanalyses and monthly surface climatological datasets. *Quarterly Journal of the Royal Meteorological Society*, **in press**, doi: 10.1002/qj.2949.

Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. M. B. Tignor, and H. L. Miller, Eds., 2007: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK.

Steinhilber, F., J. Beer, and C. Fröhlich, 2009: Total solar irradiance during the holocene. *Geophysical Research Letters*, **36 (19)**, doi:10.1029/2009GL040142.

Stevens, B., 2013: Aerosols: Uncertain then, irrelevant now. *Nature*, **503**, 47–48, doi:10.1038/503047a.

Stoffel, M., and Coauthors, 2015: Estimates of volcanic-induced cooling in the northern hemisphere over the past 1,500 years. *Nature Geoscience*, **8**, 784 – 788, doi:10.1038/ngeo2526.

Suckling, E., E. Hawkins, J. Eden, and G. J. van Oldenborgh, 2016: An empirical model for probabilistic decadal prediction: global attribution and regional hindcasts. *Climate Dynamics*, in press, doi:10.1007/s00382-016-3255-8.

Sutton, R., E. Suckling, and E. Hawkins, 2015: What does global temperature tell us about local climate? *Phil. Trans. A*, **373**, 20140 426, doi:10.1098/rsta.2014.0426.

Thordarson, T., and S. Self, 2003: Atmospheric and environmental effects of the 1783–1784 laki eruption: a review and reassessment. *Journal of Geophysical Research: Atmospheres*, **108 (D1)**, 1–29.

Usoskin, I. G., and Coauthors, 2015: The maunder minimum (1645–1715) was indeed a grand minimum: A reassessment of multiple datasets. *Astronomy & Astrophysics*, **581**, A95.

Van Engelen, A., J. Buisman, and F. Ijnsen, 2001: *A millennium of weather, winds and water in the low countries*. Springer.

Van Engelen, A., and J. Nellestijn, 1990: Monthly, seasonal, and annual means of the air temperature in tenths of centigrade in de bilt, netherlands, 1706–1990. KNMI, De Bilt.

Vinther, B. M., P. Jones, K. Briffa, H. Clausen, K. Andersen, D. Dahl-Jensen, and S. Johnsen, 2010: Climatic signals in multiple highly resolved stable isotope records from greenland. *Quaternary Science Reviews*, **29 (3)**, 522–538.

Widmann, M., H. Goosse, G. van der Schrier, R. Schnur, and J. Barkmeijer, 2010: Using data assimilation to study extratropical northern hemisphere climate over the last millennium. *Climate of the Past*, **6 (5)**, 627–644, doi:10.5194/cp-6-627-2010.

Zanchettin, D., C. Timmreck, O. Bothe, S. J. Lorenz, G. Hegerl, H.-F. Graf, J. Luterbacher, and J. H. Jungclaus, 2013: Delayed winter warming: A robust decadal response to strong tropical volcanic eruptions? *Geophys. Res. Lett.*, **40 (1)**, 204–209, doi:10.1029/2012GL054403.

LIST OF FIGURES

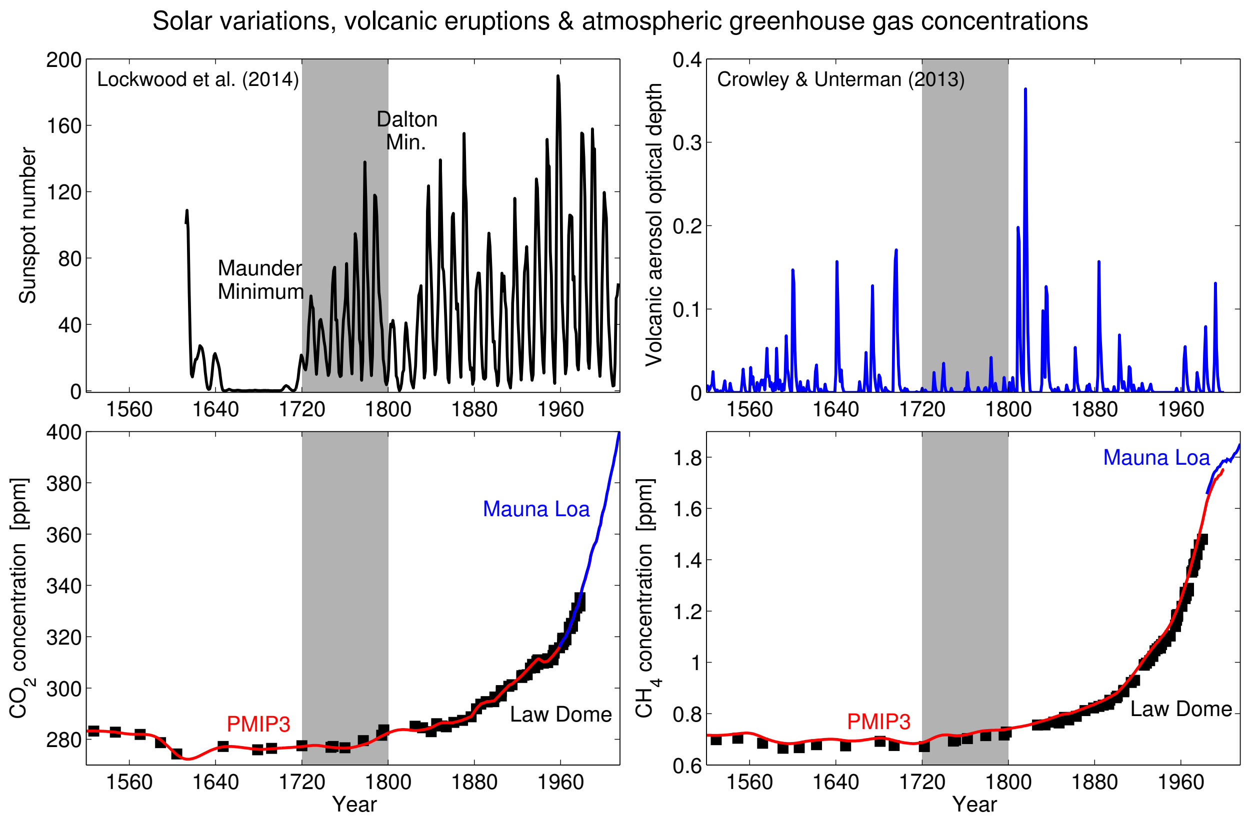


FIG. 1. Historical natural forcings and greenhouse gas variations. Top left: annual sunspot number since 1612, with the Maunder Minimum and Dalton Minimum indicated (Lockwood et al., 2014). Top right: estimated global volcanic aerosol optical depth (Crowley and Unterman, 2013). Bottom row: the Law Dome greenhouse gas data (MacFarling Meure et al. (2006), black) for carbon dioxide (left) and methane (right), along with the annual means from Mauna Loa (Keeling et al. (2001), blue) and PMIP3 assumed values (Schmidt et al. (2012), red). Note there is a 16ppb offset applied to the smoothed Law Dome methane concentrations to produce a global mean as used by PMIP3 to account for the interhemispheric gradient. The 1720-1800 period is denoted by the grey shaded region in all panels.

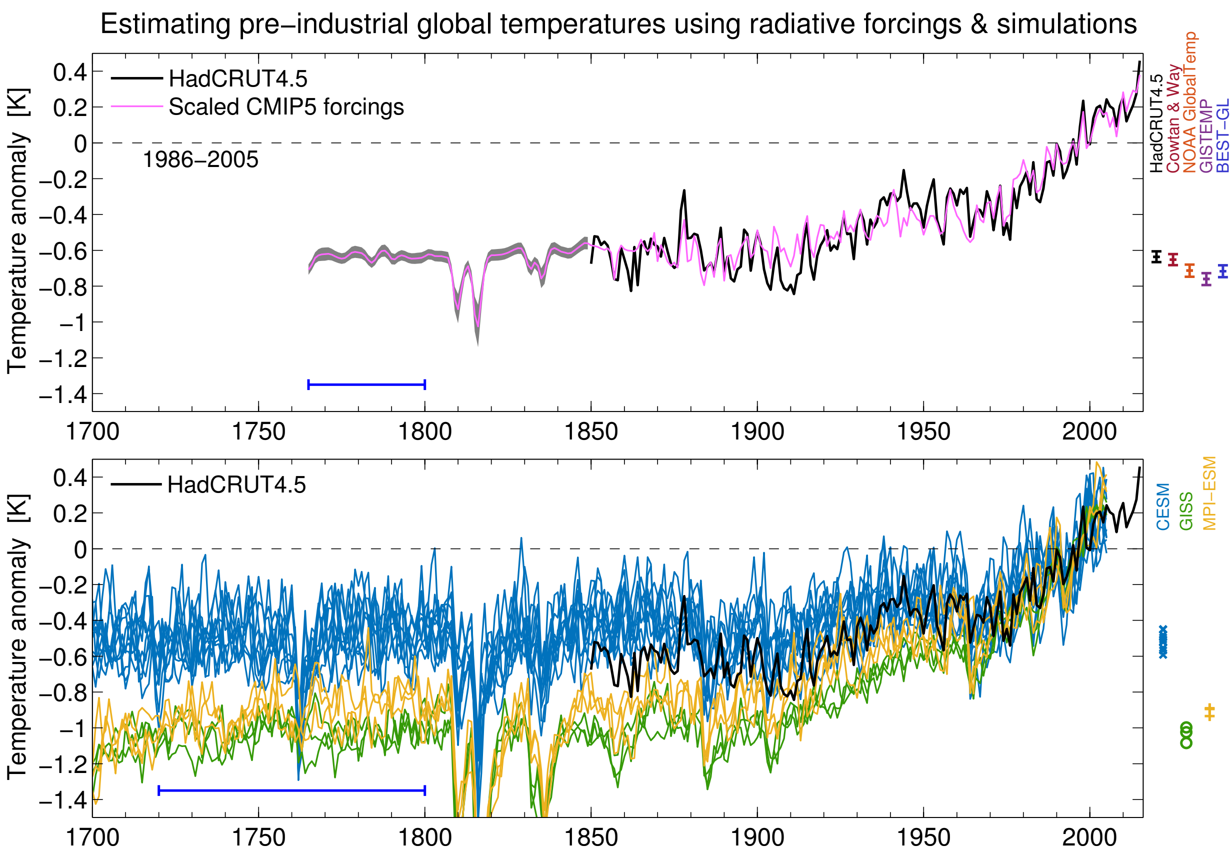
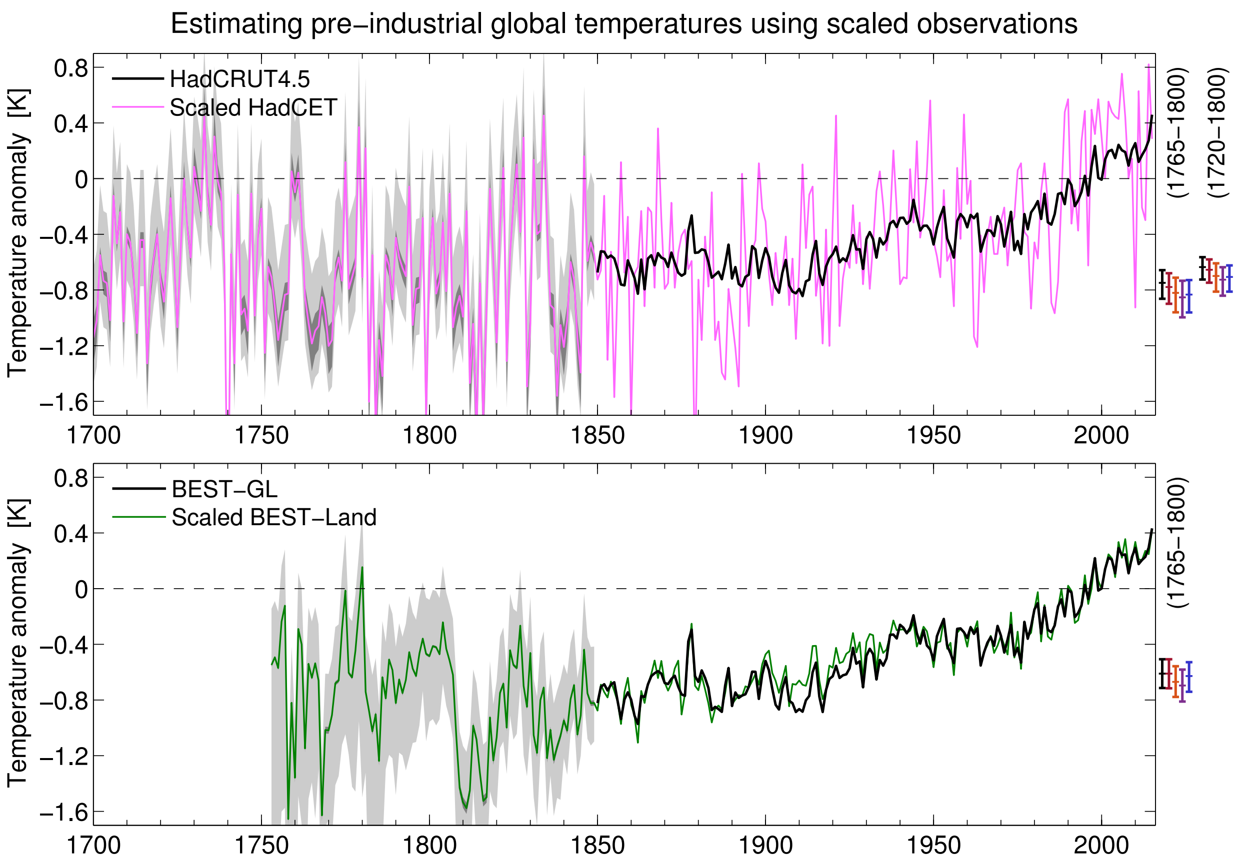
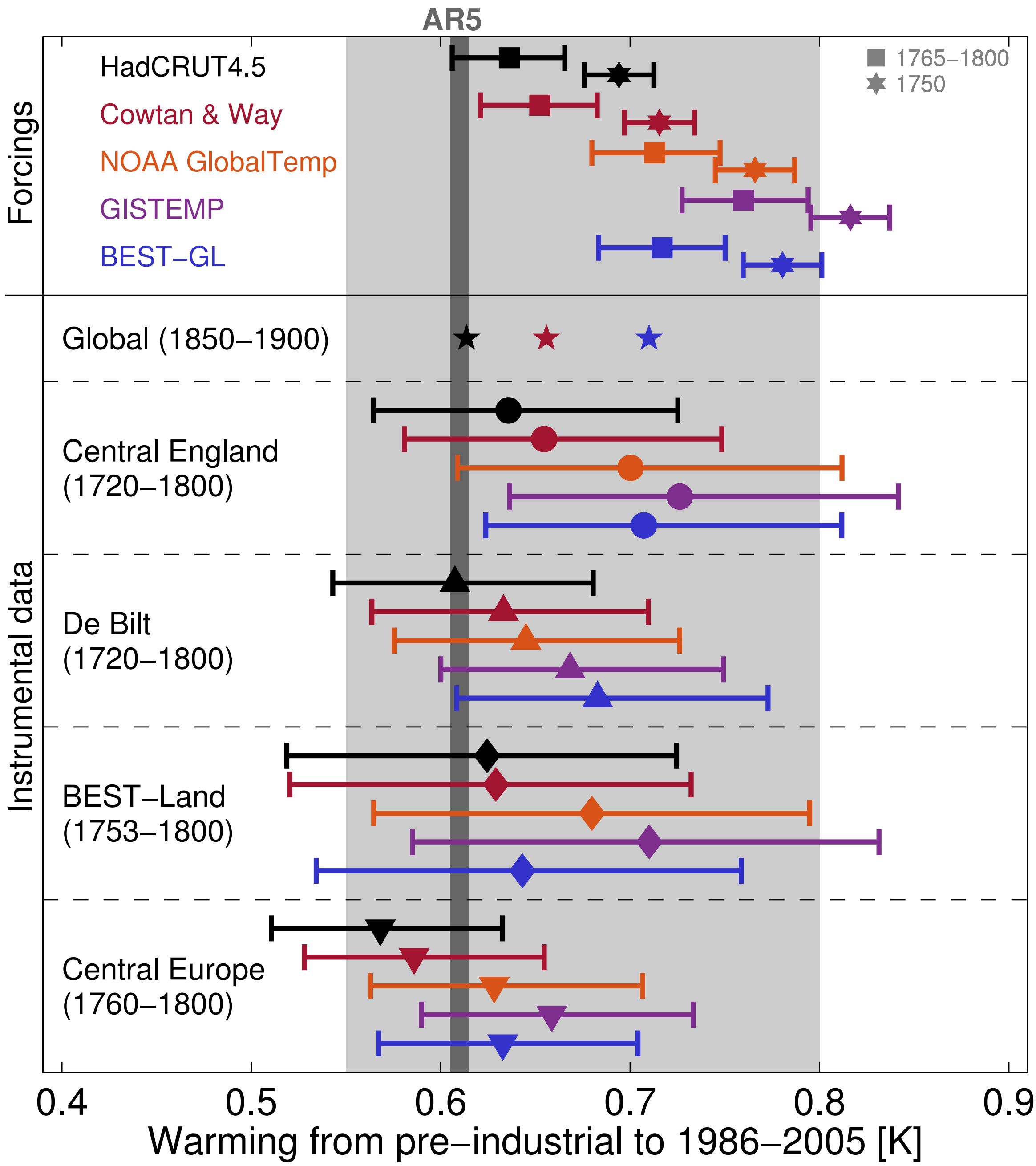


FIG. 2. Top panel: estimating global pre-industrial temperature using scaled radiative forcings (pink), using HadCRUT4 (black) as the reference. The grey shading represents the uncertainty in the regression. Estimated global temperature anomalies for 1765-1800 are given for all five global temperature datasets (right hand side, as labelled). Bottom panel: simulated global temperature anomalies in the Last Millennium Ensembles (LMEs) and estimates for the change since 1720-1800 for the range of ensemble members of CESM (blue), GISS (green) and MPI-ESM (orange). In both panels the blue horizontal bars indicate the period used for averaging. The 1986-2005 reference period is represented by the black dashed line.



825 FIG. 3. Estimating global pre-industrial temperature using scaled annual mean observations for CET scaled to HadCRUT4 (top) and BEST-Land scaled to BEST-GL (bottom), relative to 1986-2005 (dashed black). The dark grey shading (hardly visible) represents the uncertainty in the regressions and the light grey shading the uncertainty in the observations. The sets of five error bars on the right hand side use the different global temperature datasets, with the same ordering as in the top panel of Fig. 2, for the two different averaging periods as labelled. Note the vertical scale is different from Fig. 2.



831 FIG. 4. Summarising the evidence for annual mean global temperature change from pre-industrial until 1986-
832 2005 using each dataset. The horizontal bars represent the 5-95% uncertainty ranges for the different sources of
833 evidence. Results for the radiative forcing approach are shown averaged over 1765-1800, and for 1750. The top
834 row in the instrumental observations section shows the observed change since 1850-1900 (where available). For
835 the instrumental data the longest timeseries during the pre-industrial period are used: CET and De Bilt (1720-
836 1800), BEST-Land (1753-1800) and Central Europe (1760-1800). The light grey shading shows the assessed
837 likely range and the dark grey line indicates the IPCC AR5 assessment (0.61°C, Kirtman et al., 2013).

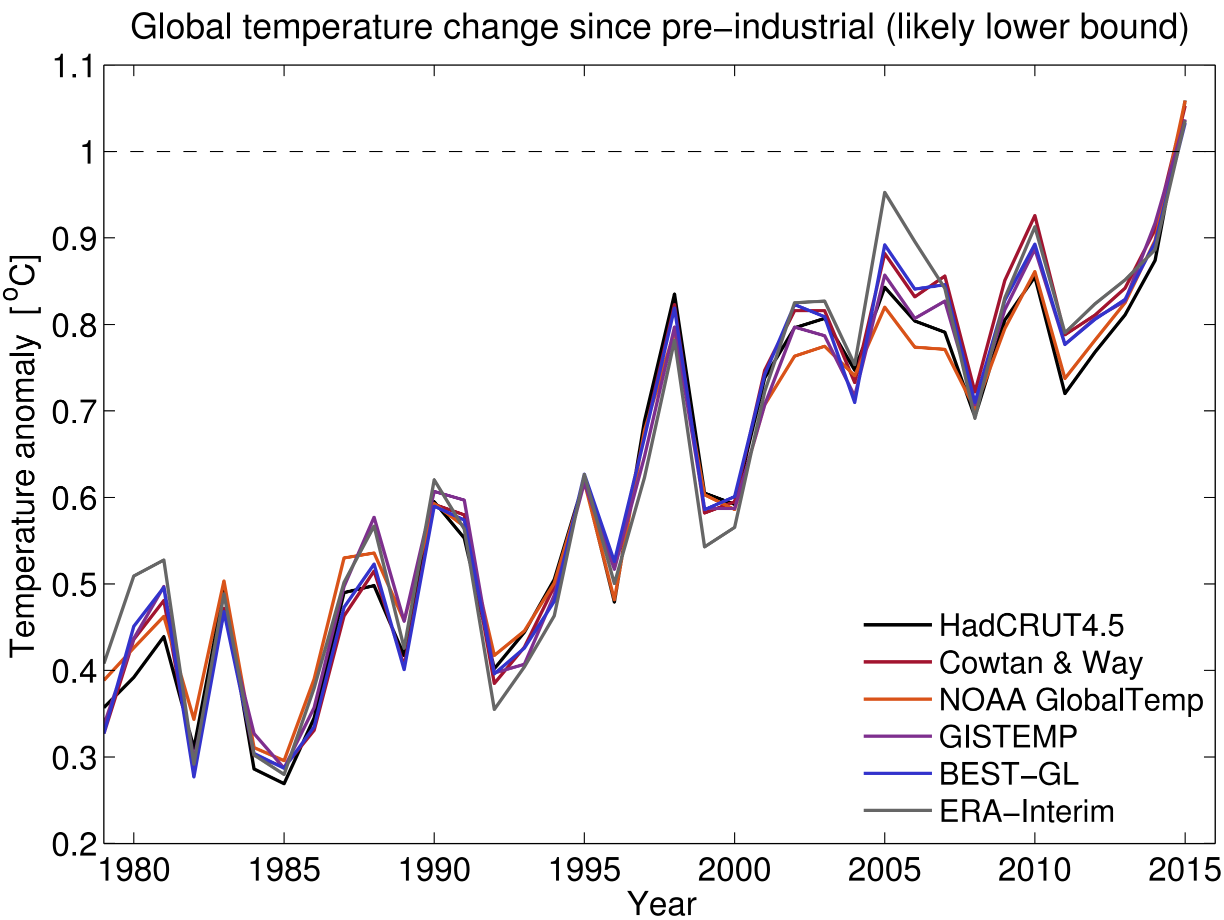
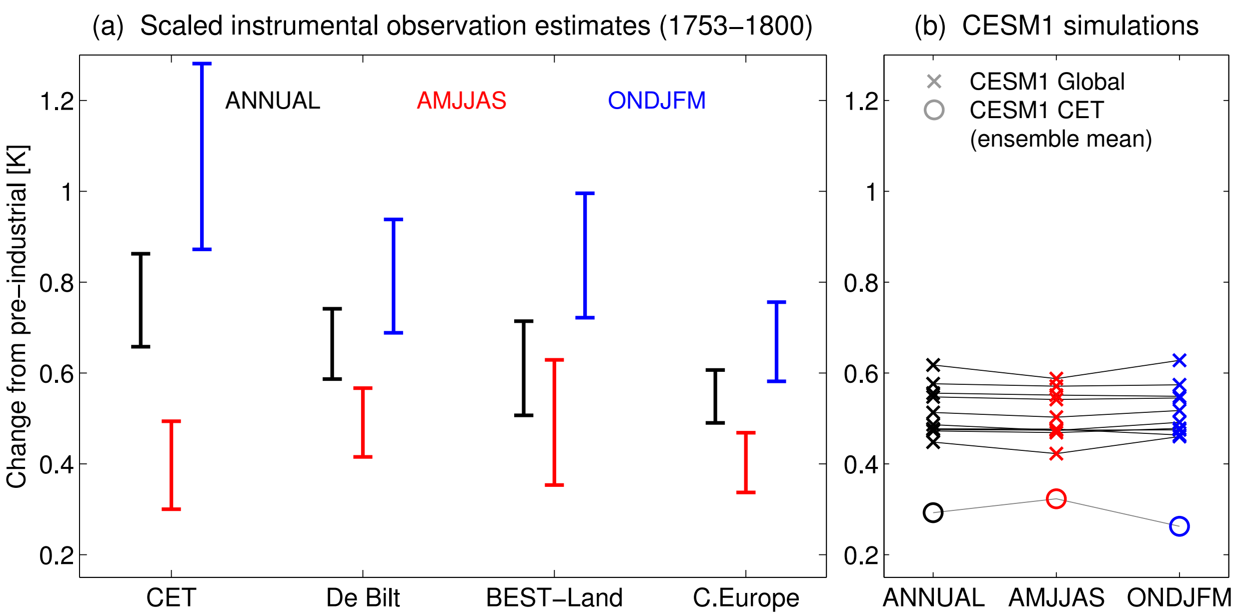


FIG. 5. Global mean temperature relative to pre-industrial in six datasets, using the likely lower bound (0.60°C) for warming from pre-industrial to 1986-2005. The change in the ERA-Interim reanalysis (Dee et al., 2011) relative to 1986-2005 is included with the five global temperature datasets discussed. The 1996-2015 period is $0.16-0.19$°C warmer than 1986-2005.



842 FIG. 6. Seasonal differences in warming rates. (a) Derived scaled warming from 1753-1800 to 1986-2005 (using Approach 3) for annual means (black) and for the extended seasons (April to September - AMJJAS, red, and October to March - ONDJFM, blue) for the different regional timeseries, all using annual mean HadCRUT4 as the reference dataset. (b) Seasonal warming derived from the CESM1 LME simulations for the global mean (crosses, with black lines linking the same ensemble members in each season) and for the ensemble mean of simulated CET (circles).